

Türkiye’de Online Kitap Satışlarının Veri Madenciliği Yöntemiyle Analizi

DOI: 10.26466/opus.793764

*

Meltem Ulusan Polat * – Mehmetcan Babaoğlu **

* Dr. Öğretim Üyesi, Marmara Üniversitesi

E-Posta: meltem.ulusan@marmara.edu.tr

ORCID: [0000-0002-0673-5185](https://orcid.org/0000-0002-0673-5185)

** Finansal Analist, Mercedes – Benz Otomotiv

E-Posta: mehmetcanbabaoglu@gmail.com

ORCID: [0000-0001-8911-5936](https://orcid.org/0000-0001-8911-5936)

Öz

Online satış ağları her geçen gün hayatımızda daha büyük bir yer kaplamaktadır. Günümüzde, rekabentin artması ve insanların geniş bir ürün yelpazesine tek bir tuşla ulaşabilme rahatlığı kitap satışları açısından da online satışı ön plana çıkartmıştır. İçinden geçmekte olduğumuz pandemi sürecinde de online satış en çok gelişen ve kârlılığı yükselen alanlardan biri olmuştur. Türkiye’de okuma alışkanlığının artması ve okuyucu tercihlerine yönelik doğru stratejilerin geliştirilmesi açısından çalışmanın önemi büyüktür. Bu çalışmada, popüler online satış sitelerinin 2014-2019 yılları arası en çok satan kitaplarından derlenen 508 adet kitap verisi isim, yazar, tür ve yayınevi parametreleri bazında RapidMiner programı ile veri madenciliği kapsamında K-Means algoritması kullanılarak kümeleme analizine tabi tutulmuştur. Adı geçen parametreler kapsamında özellikle yerli ve yabancı romanların üstünlüğü ortaya çıkarken bazı tematik kitap türlerinde ise günden güne bir gerileme yaşandığı gözlemlenmiştir. Ulaşılan analiz sonuçları ve bulgular özellikle fikir aşamasındaki yazar ve maliyetleri üstlenecek olan yayınevlerinin tercihlerinde yol gösterici olmaya yönelik detaylı olarak değerlendirilmiştir.

Anahtar Kelimeler: veri madenciliği, kümeleme, K-Means, online kitap satış

The Analysis of Online Book Selling in Turkey by Data Mining

*

Abstract

Online sales networks occupy a greater space in our lives, day by day. The increasing competition and access to wide product range with just one click bring online book sales to the forefront. During current pandemic, online sales has been one of the top areas in terms of profitability. The research has a major importance in terms of increasing reading habit in Turkey and developing accurate strategies due to reader preferences. In the context of this research, best-selling books data between 2014-2019, obtained from popular web-sites, involving parameters such as name, author, genre and publisher of 508 books have been analyzed by data mining method (clustering – K Means algorithm) using RapidMiner. The findings within the scope of the parameters indicated the superiority of both native Turkish and foreign novels, while some thematic books declined. The results have been deeply evaluated particularly to guide authors at the idea stage and to determine a road map for publishers while bearing the costs.

Keywords: *data mining, clustering, K-Means, online book selling*

Giriş

Günümüzde veri madenciliği hayatın her alanında büyük önem kazanmıştır. Veri yığınları içerisinde işe yarayacak olan faydalı bilgiyi arındırmak ve analiz neticesinde daha anlamlı bilgi elde ederek bunu yorumlamak, veri madenciliğini ön plana çıkartmıştır. Bu tarz bir analiz salt insan yeteneğini aşmakta olup, özellikle dijital veri miktarındaki artışa karşılık araştırma ve uygulama yapan kişi sayısının değişmemesi, alanın gelişmesindeki önemli bir diğer etken olmuştur. Bu gereksinimlerin sonucunda büyük miktarda verinin işlenmesinin yanı sıra global ve rekabetçi bir ortamda doğru karar verebilme ihtiyacı da veri madenciliğini kaçınılmaz kılmıştır (Savaş ve diğerleri, 2012).

Hayatın içerisinde her geçen gün meydana gelen gelişmeler, ticaret ve iş akışı süreçlerini de etkilemekte, veri madenciliği yoluyla elde edilen bilgiler teknoloji yardımıyla karar süreçlerine entegre edilmektedir. Geçmiş yüzyıllara dayanan ticaretin temelini oluşturan ürün seçiminden ödemeye oradan tedarike kadar olan bu süreçler, mevcut sınırları geçersiz kılan bir şekilde, elektronik ortama taşınmış ve bugün elektronik ticaret adını verdiğimiz kavramın oluşmasına ve içerik kazanmasına neden olmuştur (Özgür, 2012).

Elektronik ticaretin öncelikli olarak hâkim olduğu sektörlerden biri de tüketici taleplerine hızlı ve farklı satış kanalları üzerinden hizmet veren perakende sektörüdür. Gelişen rekabet koşulları perakendecilerin fiziksel mağazaların yanı sıra çevrimiçi mağaza ve özel alışveriş sitelerinde gibi diğer pazarlama kanallarında da aktif olmalarını zorunlu kılmıştır (Börühan ve diğerleri, 2017). Bu çalışmada, perakende sektörü kapsamında, internet üzerinden gerçekleşen kitap satışları veri madenciliği kullanılarak incelenmiştir. Yakın geçmişe kadar kitap satın alabilmek için bir büyük şehirde yaşamak ya da bu şehirlerle bağlantılı olmak bir zorunluluk iken, teknolojik altyapıyla birlikte kitap satışlarının e-ticaret kapsamında yapılabilmesi tüketici açısından süreci kolaylaştırıcı bir hizmeti beraberinde getirmiştir.

Çalışma, kayda değer satış rakamlarına ulaşan kitap satış sitelerinden temin edilen verinin veri madenciliği yöntemiyle analize tabi tutulması bağlamında, özellikle ulusal akademik yazında ciddi bir boşluğu doldurmayı hedeflemektedir. Elde edilen sonuçların gerek yayınevleri gerekse yazarlar açısından hem finansal açıdan hem de okuyucuların ilgi alanlarının belirlenme-

sinde yol gösterici olması beklenmektedir. Tüm bu çalışmaların ışığında, ayrıca, Türkiye’de okuma alışkanlığının geliştirilmesinde de bir katma değer sağlamak hedeflenmektedir.

Çalışmanın bir sonraki bölümünde verilecek olan literatür araştırmasını takiben detaylı bir metodoloji ile bulgulara yönelik sonuçlara değinilecektir.

Literatür Taraması

Veri madenciliği yöntemi kullanılarak kitap satışlarının incelendiği çalışmalar üzerine bir araştırma yapıldığında, genellikle, kitap satışlarının perakende sektörü içerisindeki ürünlerden biri olarak bir alt değişken şeklinde ele alındığı görülmüş ve salt kitap üzerine yapılan az sayıda çalışmaya rastlanmıştır. Emel ve Taşkın (2005), çalışmalarında, hediyelik eşya, müzik, cep telefon ve aksesuarları, kitap-dergi satan perakendeci bir işletmede karar ağacı yöntemiyle işletmenin gerçekleşen toplam satış hasılatının ürünlere göre dağılımının ayrıntılarını ve ürünlerin toplam satış değerlerinin toplam satış hasılatı üzerindeki göreceli önemini, müşteriler ile ilişkilendirerek bir tahmin modeli oluşturmuşlardır. Elde ettikleri sonuçlar, hangi gelir grubundaki müşterilerin hangi ürünleri ne oranda satın aldıklarını ortaya koymuş ve bu bulgulardan yola çıkarak işletmenin çok daha etkin pazarlama stratejileri gerçekleştirebileceğinin altını çizmişlerdir.

Alandaki az sayıdaki akademik çalışmadan bir tanesi, Kalikov (2006) tarafından, bir yayınevinin internet sitesindeki veri dikkate alınarak, veri madenciliği birliktelik kuralları metoduyla sepet ve sipariş tablolarının incelenmesi yoluyla yapılmıştır. Çalışma kapsamında, hangi ürünlerin kategorisinin değiştirilmesi gerektiği, kullanıcıların meslek ve ilgi alanı dağılımları, müşterilerin ilgi alanlarına göre satış grafikleri ve kullanıcıların ödeme seçenekleri ile ilgili bir analiz gerçekleştirilmiştir. Uluslararası yazında ise, Nettleton ve Baeza-Yates (2013)’e ait bir çalışmada, grafik ölçümleri ve algoritmalar kullanılarak bir elektronik kitap satış sitesi bağlamında kullanıcı, kitap ve yayınevi ilişkileri analiz edilmiştir. Bulgulardan yola çıkılarak çapraz satış ve kitap kataloglama açısından şirkete faydalı bilgiler sağlamak hedeflenmiştir.

Son yıllarda yapılan kayda değer bir diğer çalışma da Sharma ve diğerleri (2019) tarafından gerçekleştirilmiştir. Çalışmada, regresyon, karar ağacı ve yapay sinir ağları kullanılarak çeşitli faktörler kapsamında Amazon sitesin-

deki kitap satışlarını tahminlemede en etkili modelin hangisi olduğu araştırılmış ve her üç model için de inceleme, negatif duyarlılık, pozitif duyarlılık ve bunların birbirleriyle etkileşimlerinin satış tahmininde öne çıkan faktörler oldukları tespit edilmiştir. Bu bulguların online satışçılar açısından satış hacmi ve tedarik zinciri stratejilerinin etkin belirlenmesinde faydalı olacağı öngörülmüştür.

Teknolojik altyapının gelişimi ve veri madenciliğinin öne çıkması bunun yanı sıra Türkiye açısından kitap okumanın önemine vurgu yapılması saikleriyle bu çalışmanın motivasyonu geliştirilmiş ve elde edilecek bulguların başta online kitap satış siteleri olmak üzere yayıncılar, yazarlar ve okuyuculara yol göstermesi hedeflenmiştir. Ayrıca, dünya genelinde incelendiğinde, en büyük cirolara sahip alışveriş siteleri kitap satış siteleridir (Özgür, 2012). Bu referans noktalarından hareketle, Türkiye Yayıncılar Birliği, 2019 yılı için hazırladığı kitap pazarına yönelik raporunda, Türkiye kitap perakende pazarında bir önceki yıla göre TL bazında yaklaşık %27'lik bir artış olduğunu belirtmiştir. Başta dolar kurunda yaşanan yükseliş olmak üzere, fiyatlardaki bu yükselişte temel ekonomik göstergelerle doğrudan bir ilişki söz konusudur. Okuyucuya maliyet olarak yansıyan bu durumun iyileştirilmesinde, okuyucu tercihlerine göre satışa sunulan kitaplar ve buna bağlı oluşacak bir fiyatlama politikası önem arz etmektedir.

T.C. Kültür ve Turizm Bakanlığı'nın yakın geçmişte hazırladığı Türkiye Okuma Kültürü Haritası isimli çalışmada ise, veriler kapalı ve açık uçlu sorulardan oluşan "anket tekniği" aracılığı ile toplanmıştır. Alan çalışmasında elde edilen anketlerin kontrolü yapıldıktan sonra, veri kodlama ve yükleme işlemleri tamamlanmış, analizleri yapılmış ve raporlaştırılmıştır. Raporda, yılda ortalama 7.2 adet kitap okunduğu açıklanmıştır. Katılımcıların %84.2'si düzenli olarak takip edilen bir yazarları olmadığını açıklamıştır. Ankete göre, düzenli olarak okunan üç yazar ise Ömer Seyfettin, Ayşe Kulin ve Orhan Pamuk olarak belirlenmiştir. Kitap seçiminde katılımcıların %90.16 oranında yayınevi tercihi yapmadığı ortaya çıkmıştır. Tercihler arasında ise en çok tercih edilen yayınevi Can Yayınları'dır (Türkiye Okuma Kültürü Haritası, 2011).

Özgür (2012) tarafından yapılan bir diğer çalışmada ise, kitapların türlerine göre satın alınma oranlarında da en başta % 40,08'le edebi eserlerin gelmekte olduğu ortaya konmuştur. Akademik yayınların ise % 2,42 ile en sonda olduğu görülmüştür. Türkiye için yapılan bu çalışmalara bakıldığında,

okuma alışkanlığının geliştirilmesinden hareketle, okuyucu, yayınevi, yazar ve kategori açısından veriyi ayrıştırma ve analiz ederek taraflara yönelik doğru politikalar üretmenin önemi bir kez daha ortaya çıkmaktadır.

Amerika Birleşik Devletleri’nde “EPJ Data Science” isimli akademik dergide yayımlanan bir çalışmada, en çok satan kitapların, satış zirvesine on haftadan daha kısa sürede ulaştığı ortaya çıkmıştır. Çalışmada 4500 adet kitap ve satış verisi kullanılmıştır. Modelleme olarak ise yazarlar kendi modellerini oluşturmuşlardır (Yücesoy ve diğerleri, 2018).

Birleşik Krallık’ta 2018’de satılan kitap verilerinden yola çıkarak yapılan bir çalışmada ise, en çok satılan kitap türünün roman olduğu, çocuk kitaplarının da %31 ile tahmin edilenden çok daha fazla satıldığı tespit edilmiştir (Rowe, 2019).

Veri madenciliği yöntemiyle kitap satışlarının analiz edileceği bu çalışmanın, daha önce de belirtildiği üzere, özellikle metodik anlamda Türk akademik yazınına bir katma değer sağlayacağı ve alandaki öncü çalışmalardan biri olabileceği düşünülmektedir.

Metodoloji

Çalışmanın bu bölümünde amaç ve kapsam ile çalışmada kullanılan yöntem detaylı olarak açıklanacaktır.

Çalışmanın Amacı ve Kapsamı

Bu çalışmada, 2014-2019 yılları arasında en güvenilir olarak değerlendirilen (Kara, 2018) üç adet online kitap satış sitesinden temin edilen “en çok satan kitap, çok satan yazar, kitap türleri, en çok satan kitapların yayınevleri” parametrelerinden oluşan veri seti veri madenciliği yöntemiyle analiz edilecektir.

Çalışma kapsamında yukarıda adı geçen üç sitenin seçilmesinin nedeni, güvenilir olmalarının yanı sıra, geriye dönük satış verisine ulaşılabilmesi ve bu sitelerin tüketici açısından da oldukça popüler online kitap satış siteleri olmalarıdır. Özellikle veri temini açısından sadece bu üç sitenin mercek altına alınıyor olması aynı zamanda çalışmanın bir kısıtı olarak da değerlendirilebilir.

Bu inceleme sırasında 508 adet kitap bilgisi üzerinden analizler gerçekleştirilmiştir. Çalışmanın sonucunda yıl bazlı çok tercih edilen yazar, yayınevi,

kitap ve kitap türlerinin sonuçları elde edilmiştir. Bu sonuçlardan yola çıkılarak, toplumdaki okuma tercihleri hakkında bir fikir edinmek ve bu yolla yayınevlerinin basım ve satış politikalarını, gerek cirolarını gerekse karlılıklarını arttırmaya yönelik, belirleyebilmeleri açısından yol gösterici olmak amaçlanmaktadır. Ayrıca, bu bilgiler ışığında, tedarik zinciri açısından da yayınevleri açısından etkin politikaların üretilmesi mümkün olabilecektir.

Çalışmanın Yöntemi

Çalışmada, veri madenciliği yöntemlerinden biri olan kümeleme yöntemi kullanılmıştır. Daha önce de belirtildiği üzere, geleneksel istatistik yöntemlerle büyük boyuttaki veriyi çözümlmek kolay olmadığından, veri madenciliği gibi verileri işlemek ve çözümlmek için özel yöntemlere gereksinim duyulmuştur (Özkan, 2008).

Veri madenciliği modelleri, işlevleri açısından, 3 grupta toplanır;

1. Sınıflandırma (Classification) ve Regresyon (Regression),
2. Kümeleme (Clustering),
3. Birliktelik Kuralları (Association Rules)

Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları modelleri ise tanımlayıcı modellerdir (Özekes, 2003). Çalışmada, kümeleme kapsamında, RapidMiner programındaki kümeleme modelinin K-Means (K-ortalama) algoritması kullanılmıştır. Analize geçmeden önce, metotla ilgili önemli kavramsal bilgiler aşağıda yer almaktadır.

Kümeleme: Bir denetimsiz öğrenme metodu olan kümeleme, birbirine benzer verileri sınıflara ayırarak verileri özetleme veya fiziksel olarak gruplandırma sürecidir. Kümeleme yöntemi, büyük veri kümelerinin anlamlı şekilde parçalanarak alt kümelerine ayrıldığı ve benzer grupların bir arada ele alındığı bir süreçtir (Lin ve diğerleri, 2005).

Kümeleme, veri kümelerinde verileri birbirinden ayıran başka bir yöntem olan diskriminant analizinden ayrılmaktadır. Kümelemede küme sayısı bilinmemekte ancak analiz sonucundaki veriden elde edilmektedir. Bununla birlikte kümelemede herhangi bir fonksiyon elde edilip sonrasında diğer veriler için kullanılma durumu yoktur; ayırma işlemi sadece o verilerin özellikleri kullanılarak yapılır.

Kümeleme süreci, her bir verinin farklı özelliklerini ifade eden değerlerinin, diğer bir veriden farklı veya aynı olmasına göre işler. Farklılık ya da ayınlık kümeleme algoritmaları için uzaklıkları ifade eder. Kümeleme işleminde küme içindeki elemanların benzerliği fazla, kümeler arası benzerlik ise az olmalıdır (Sarman, 2011).

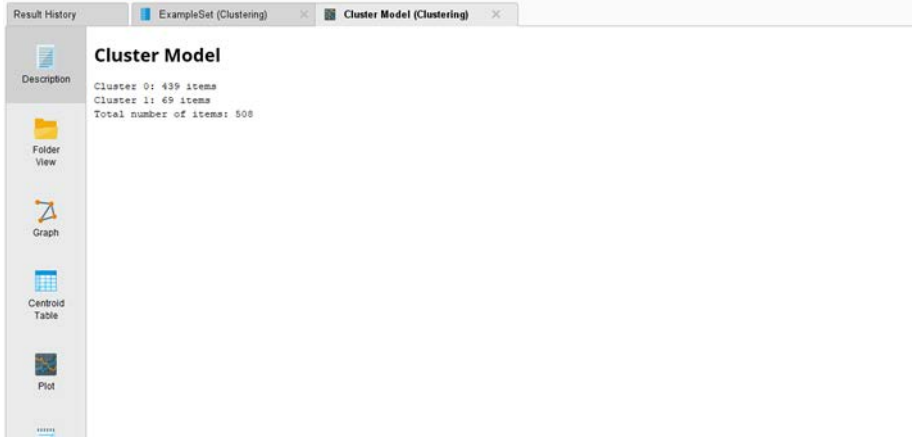
Kümeleme yöntemleri, farklı sektörlerde birçok amaçla kullanılmaktadır. En sık rastlanan kullanımlarından biri, pazarlama sektöründe müşterilerin segmentlere ayrılarak farklı segmentler için farklı pazarlama stratejilerinin geliştirildiği durumlardır (Çelik, 2009). Bir diğer yaygın ve oldukça önemli kullanım alanı olarak, finansal performansa yönelik araştırmalar gösterilebilir. Bankacılık sektöründe müşterilere yönelik yapılan kredi performans analizleri ile halka açık şirketlere yönelik finansal performans analizleri (Arı ve diğerleri, 2016) bu kategori içerisinde sayılabilir.

K-Means Algoritması: K-Means algoritması bir kümeleme algoritmasıdır. Algoritma, istatistiksel olarak benzer nitelikteki kayıtları aynı gruba sokar. Bir elemanın yalnızca bir kümeye ait olmasına izin verilir. Küme merkezi kümeyi temsil eden değerdir (Akça, 2014). Bu çalışmada, veri yıl bazlı olduğundan ayrıca benzer nesnelere birlikte gruplayarak olabildiğince ayrıntılı açıklayabilmek amacıyla K-Means algoritması kullanılmıştır.

RapidMiner: RapidMiner makina öğrenmesi ve veri madenciliği ihtiyaçlarına yönelik olarak yazılımla aynı ada sahip firma tarafından geliştirilen bir yazılım platformudur (Aksan, 2019). Yazılım çoğunlukla iş ve ticari amaçlarda kullanılmasının yanı sıra eğitim, araştırma, hızlı prototipleme ve uygulama geliştirme gibi farklı amaçlarda da kullanılır. Ayrıca, veri madenciliği sürecinin bütün aşamaları RapidMiner tarafından desteklenmektedir. Bu nedenle, veri hazırlama, sonuçları görselleştirme, doğrulama ve optimizasyon gibi amaçlarla kullanılabilir (Yıldız ve Şeker, 2016).

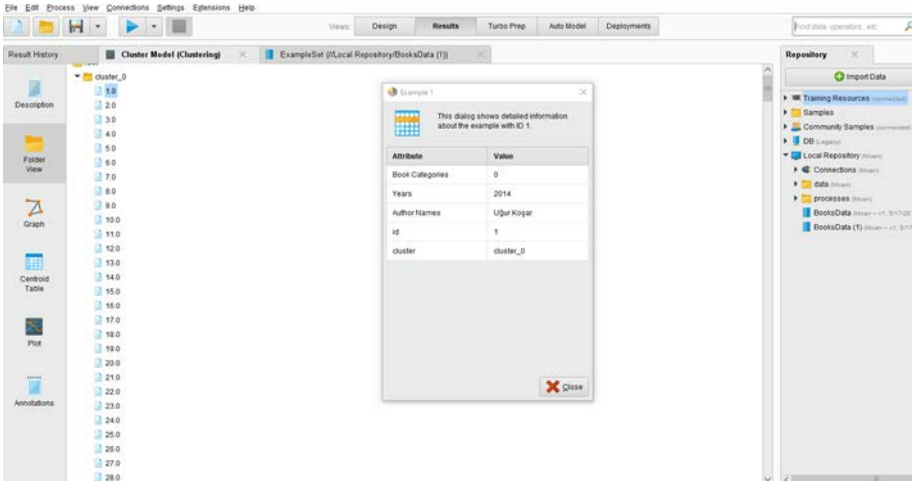
RapidMiner İle Kümeleme Analizi Uygulaması

Çalışmanın bu aşamasında, adım adım yazar ve kitap türleri bazında kitap satışlarına yönelik analizin ilerleme süreci gösterilmektedir. Uygulanan yöntemin ve çıktıların daha net ortaya konulabilmesi açısından süreç ekran çıktılarıyla desteklenmektedir.



Şekil 1. RapidMiner-Analiz için hazırlanmış iki adet kümeyi yansıtan ekran görüntüsü

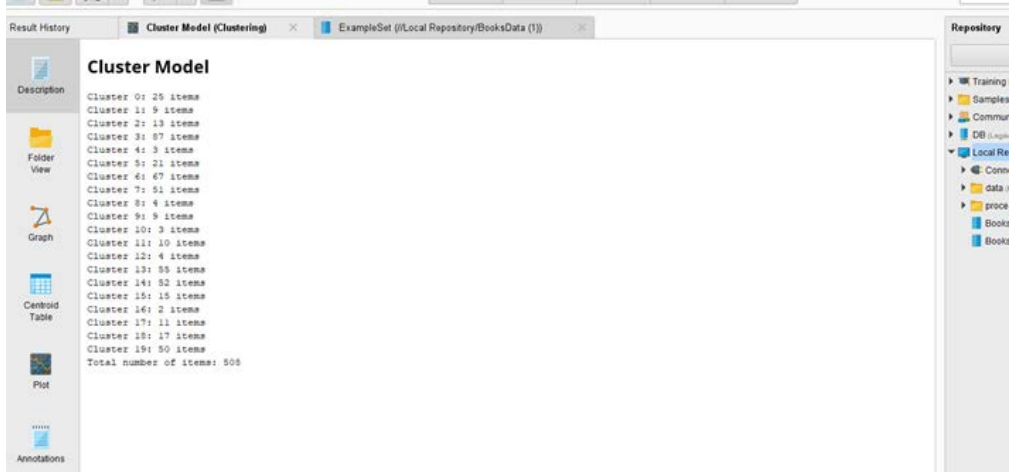
Şekil 1’de görüldüğü üzere, “Results” (Sonuçlar) penceresinde “Description” (Tanımlama) sekmesinde veri setimizdeki yazar adları ve kitap türlerine göre iki küme oluşmuştur. Birinci kümede 439 öge, ikinci kümede ise 69 ögenin bulunduğu görülmektedir. Burada kümelerin numaralandırmaları, ekran görüntüsünde de görüldüğü gibi, 0’dan başlamaktadır.



Şekil 2. RapidMiner - Folder View ile küme elemanlarının görüntülediği ekran görüntüsü

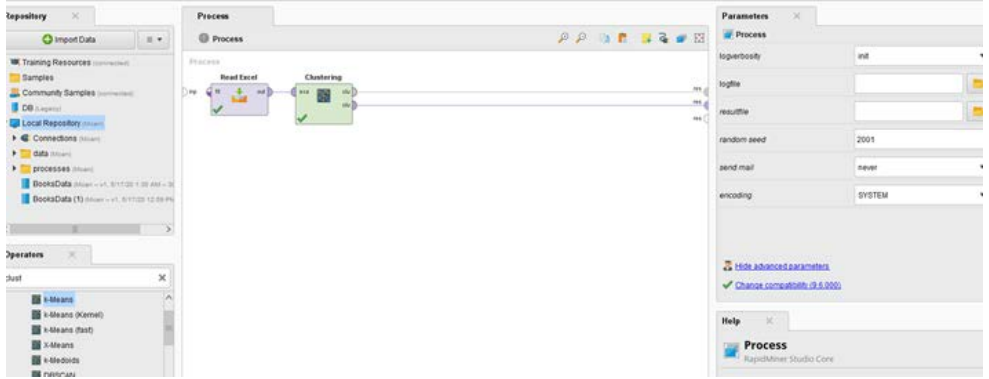
Bir sonraki aşamada “Folder View” (Dosya Görünümü) sekmesinde “Id” numarası 1 ile ifade edilmek istenen, söz konusu verilerin ve değerlerin 1 numaralı gözleme ait olduğudur. Örnek ekran görüntüsünde de görüldüğü üzere, 1 numarada yazar Uğur Koşar ve kitabın yılı 2014 olarak verilmektedir. Bu şekilde tüm gözlemler listelenmektedir. Bu verilerin gözleme ait olduğu veri setine bakılarak kontrol edilebilir.

K parametresine 2 değeri verildiğinde oluşan iki kümenin öğelerine bakıldığında, birbirine yakın değerlerin bir kümede toplandığı görülmüştür. Detaylı sonuç görmek için ise k’ya 20 verildiğinde yirmi farklı küme oluşmuştur. Şekil 3’te görüldüğü gibi, kümelerin öğe sayısında daha eşit bir dağılım vardır.



Şekil 3. RapidMiner - Analiz için hazırlanmış yirmi adet kümenin ekran görüntüsü

Analizin bu aşamasında, “Design” (Tasarım) sekmesine geri dönülüp Clustering (kümeleme) operatörünün diğer “clu port”u da “Result” portuna bağlanır (Şekil 4) ve tekrar çalıştırma butonuna basılır. Clu port’un bağlanması ile verilerin sonucu bir çıktı olarak gelir ve istatistiksel olarak görüntülenebilir. Bu işlem yapıldığında “Results” sekmesinde Cluster Model (küme modeli) ile birlikte ExampleSet (örnek küme) de görüntülenir.



Şekil 4. RapidMiner - Clustering operatörünün result portuna bağlanmasına dair ekran görüntüsü

Kümeleme Analizi Sonuçlarının Yorumlanması

“ExampleSet” sekmesinde sol tarafta “Statistics” (İstatistikler) sekmesine tıklandığında verinin türleri ve kullanılan veri setinde hiç kayıp değer (missing value) olmadığı görülmektedir. Ayrıca burada verilerin en düşük ve en yüksek değerleri de gösterilmektedir. Kümelerin kaç ögesi olduğu yine bu bölümde görülmektedir (Şekil 5).

Name	Type	Missing	Statistics	Filter (5 / 5 attributes)
id	Integer	0	Min: 1, Max: 508, Average: 254.500	
Author Names	Polynomial	0	Least: Şule Gürbüz (1), Most: Ayşe Kulin (8), Average: Ayşe Kulin (8), Haruki Murakami (8), [346 more]	
cluster	Nominal	0	Least: cluster_16 (2), Most: cluster_3 (87), Average: cluster_3 (87), cluster_6 (67), [18 more]	
Book Categories	Integer	0	Min: 0, Max: 23, Average: 3.959	
Years	Integer	0	Min: 2014, Max: 2019, Average: 2016.293	

Şekil 5. RapidMiner - Veri setinin ve kümelerin istatistiksel değerlerine ait ekran görüntüsü

“Visualizations” sekmesine tıklandığında, aşağıdaki resimde görüldüğü gibi, kümelerin grafiksel olarak gösterimine ulaşılır. Burada x eksenini yıllar, y eksenini ise kitap türleri olarak seçilmiş ve yazar adlarına göre grafiğin oluşması istenmiştir. Grafikteki her bir renkli nokta ayrı bir yazar adını göstermektedir.

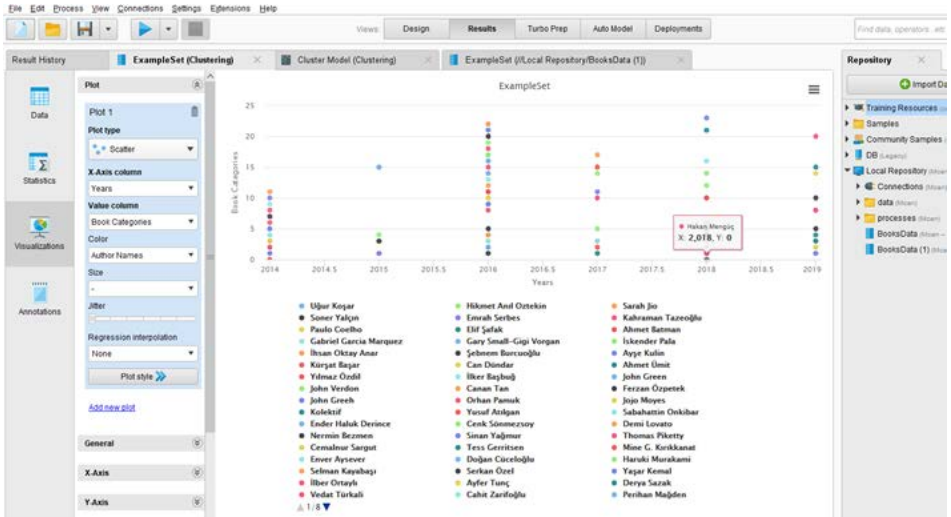


Şekil 6. RapidMiner - Kümelerin grafiksel olarak gösterimine ait ekran görüntüsü

Grafiğin sonucuna bakıldığında, yıllara göre hangi yazarın hangi kitap türüyle başarılı olduğu görülmektedir. Örneğin, 2016 yılındaki en yüksek değerin üzerine gelindiğinde Franz Kafka’nın 22 numaralı (Felsefe-Bilim) türüyle başarılı olan yani “çok satan yazar” olduğu yorumu yapılabilir. Buna karşılık, 2018 yılındaki en düşük değerin üzerine gelindiğinde, Hakan Mengüç’ün 0 numaralı (Din-Mitoloji) kitap türüyle “en az satan yazar” olduğu görülebilir. Şekil 7 ve 8’de bu örnekler görülmektedir.



Şekil 7. RapidMiner - 2016 yılında çok satan yazarın ve kitap türünün görüntülediği ekran görüntüsü



Şekil 8. RapidMiner - 2018 yılında en az satan yazar ve kitap türünün görüntülediği ekran görüntüsü

Analizi derinleştirerek farklı yorumlar yapmak için bu sefer kümelere göre x eksenini yıllar ve y eksenini ise "id" (kimlik) olarak seçilir. Buna göre oluşan grafikte; yıllardaki en yüksek değere bakıldığında çok satan yazarın, en düşük değere bakıldığında ise az satan yazarın "id" numarası ve bu numaranın hangi kümeye ait olduğu görülebilir.



Şekil 9. RapidMiner - Yıllara göre başarılı olan yazarların id ve küme bilgilerinin görüntülediği ekran görüntüsü

Şekil 9’a bakıldığında, 2018 yılındaki en yüksek değerde “id” numarası 458 ve bu “id” numarasının ait olduğu küme (cluster_14 olarak) görülmektedir. Burada 458 numaralı “id” İhsan Oktay Anar’a aittir. “Folder View” ekranında küme öğelerinin kontrolleri yapıldığında da 458 numaralı ID’nin cluster_14 dosyasında bulunduğu görülmüştür. İster küme ister yazar adları bazlı olsun çok satan veya az satan yazarların ve onların kitap türlerinin tespiti ve buna bağlı olarak yorumları örneklerde görüldüğü gibi kümeleme analizi yöntemiyle yapılabilir.

Bulgular

Yıllar bazında yapılan kümeleme analiz sonuçlarına göre, 2014 yılında yerli roman en çok satılan kitap kategorisi olmuştur. 2015 yılından itibaren ise yabancı roman, yerli romanı geçerek her sene en çok satılan kategori olmuştur. 2019 yılında yabancı romanın satış payı %36, yerli romanınki ise %26’dır.

Yabancı romanlarda klasik eserler her sene istikrarlı bir şekilde satış rakamlarında yükselen bir ivme göstermiştir. Özellikle Rus edebiyatından Dostoyevski ve Tolstoy en çok ilgi gören yazarlardır. Çehov ise 2017 yılından itibaren daha çok tercih edilmiştir. Klasikler haricinde Harry Potter, Game of Thrones ve Yüzüklerin Efendisi romanları da çok fazla sayıda satılmıştır. Ülkemizdeki genç nüfusun fantastik edebiyata olan ilgisi satış rakamlarına da

yansımıştır. Ayrıca, bu romanların filmlerinin çekilmiş olmasının ve bu sayede popülaritelerinin artmasının da satış rakamlarındaki payı büyüktür.

Yerli romanlarda en çok ilgi gören isimler Sabahattin Ali, Zülfü Livaneli ve Ahmet Ümit'tir. Sabahattin Ali, diğer yazarların aksine, hayatta olmadığından herhangi bir yeni kitabı yayımlanmamasına rağmen her sene en çok satan yazarlardan biri olmuştur. Sabahattin Ali belgeseli ve çeşitli nedenlerle kamuoyu gündemine gelen "Kürk Mantolu Madonna" eseri yazarın popülaritesini arttırmış ve bu da satış rakamlarına yansımıştır. Zülfü Livaneli ve Ahmet Ümit ise çıkardıkları kitaplar ile o sene içinde hep ilk 10'da olmuşlardır. Bu da bize bu iki başarılı yazarı düzenli olarak takip eden ve okuyan bir kitle olduğunu göstermektedir. Zülfü Livaneli'nin başta müzisyenlik olmak üzere çok yönlü kişiliği, Ahmet Ümit'in ise kitaplarında işlediği polisiye ve gizem dolu kurgular okuyuculara son derece cazip gelmektedir.

Bulgular içerisinde ulaşılan ilginç bir sonuç olarak, 2014 yılında %8 olan dini kitap satışının 2019 yılında %1'e gerilediği gösterilebilir. Türkiye Yayıncılar Birliği'nin raporu ile de teyit edilen bu azalış, bu kitapların içeriği ve okuyucu ihtiyaçlarına ne ölçüde cevap verebildiği noktasında özellikle genç kuşağı ana ekseninde tutan ayrı bir çalışma konusu olarak irdelenebilir ve bu alandaki yazarlar ile yayınevlerine yönelik yol gösterici araştırmalar gündeme gelebilir. Bir diğer enteresan bulgu ise, 2014 yılında %1 olan tarih kitaplarının satış oranının 2019 yılında %16'ya yükselmiş olmasıdır. Sosyal medyanın da yardımı ile toplum nezdinde oldukça popüler ve saygın bir isim olan İlber Ortaylı, bu kategoride en çok satış yapan yazardır. Gündemdeki pek çok tartışmaya akademik kimliğiyle tarihsel açıdan ışık tutuyor olması, gerek sosyal medya gerekse görsel medyada sıklıkla yer alıyor olması İlber Ortaylı'yı kamuoyunda hem popüler hem de güvenilir bir konuma yerleştirmektedir. Bunun dışında, son yıllarda salt tarihin dışında kişisel gelişim alanında da eserler veriyor olması Ortaylı'nın okuyucu kitlesini genişletmektedir.

Kişisel gelişim kitaplarına olan ilgi küçük oynamalar dışında yıllar boyunca %7 civarında sabit kalmıştır. Bu da bize belirli bir okuyucu grubu tarafından düzenli olarak bu türün satın alındığını göstermektedir.

Çocuk kitabı satışlarında da her geçen yıl artış görülmektedir. Bu kategorinin 2014 yılında satış oranı %2 iken, düzenli bir artışla 2019'da %10'a yükselmiştir. Çocukların her geçen gün daha fazla telefon ve tablet kullandığı yönündeki araştırmalara rağmen, satışlarda böyle bir sonuca ulaşılması kayda

değerdir. Ancak, bu oranın sadece satışları gösterdiği gerçeğinden hareketle, bu veri üzerinden çocukların kitap okuma alışkanlıklarının arttığı sonucunu çıkarmak çok doğru olmayabilir.

2014 yılından itibaren öykü kategorisinde yıllık bazda satışlar düşmüştür. Bu alanın önde gelen yazarlarından olan Ömer Seyfettin’in kitapları son yıllarda listelere girememiştir. Anton Çehov ile 2017 sonrası bir toparlanma yaşansa da, öykü satışlarının gelecek yıllarda daha da azalacağı ön görülmektedir. Türkiye’de son yıllarda sayıları artan aylık edebiyat dergileri, öykü okuyucusu için, pek çok yazarı bir araya getirmesi ve maliyetinin kitaba göre daha düşük olması nedeniyle, oldukça cazip gelmektedir. Ayrıca, romana göre daha az sayıdaki öykü kitaplarının gerek basım gerekse tanıtımı noktasında yayınevlerinin maddi kaygularla çok agresif olmaması bir diğer sebep olarak gösterilebilir.

Öyküden kaynaklanan boşluğu politika kitaplarının doldurduğu görülmektedir. Güncel konuların en çok izlenen zaman diliminde (prime-time) televizyon ekranlarına sıklıkla yansıyor olması ve siyasetin hayatın her alanında, her kesim tarafından konuşulan, tartışılan bir olgu konumunda yer alması bu artışın bir sebebi olarak gösterilebilir.

Şiir kitapları açısından bakıldığında, satışlar 2014’ten itibaren düşüşe geçmiştir. Genç nesilde şiir okuma alışkanlığı daha çok sosyal medya ve internet üzerinden gelişmektedir. Cemal Süreya, “Sevda Sözleri” isimli eseri ile bu kategorideki en başarılı şairdir.

Edebiyat alanında, Çek yazar Franz Kafka her yıl kendine okuyucu kitlesi bulmuş bir yazardır. Elif Şafak ve İskender Pala da her yıl kitapları düzenli olarak satan Türk yazarlardır. Brezilyalı yazar Paulo Coelho’nun kitapları da ülkemizde oldukça istikrarlı bir şekilde satılmaktadır. Coelho’nun “Simyacı” adlı eseri kendisini Türk okurlar açısından popüler yapan en önemli eserdir.

Yayınevleri açısından durum incelendiğinde, kitapları en çok satan iki adet yayınevi bulunmaktadır. Bunlardan biri Türkiye’nin önde gelen yazarlarından biri tarafından 80’li yıllarda kurulan bir yayınevi olurken, diğeri de bir banka bünyesinde kurulmuştur. Her ikisi de her sene hemen hemen eşit sayıda kitapla listelerde yer almışlardır. Bir tanesi özellikle yabancı roman kategorisinde başarılı olurken, diğeri ise yerli roman kategorisinde öne çıkmıştır. Yabancı romanlarda öncü olan yayınevinin dikkat çeken yazarları arasında Gabriel Garcia Marquez, George Orwell, Stefan Zweig, Jean-Paul Sartre

sayılabilirken, yerli romanların ağırlıklı olduğu diğer yayınevini yazarlarında ise Yaşar Kemal, Tezer Özlü, Orhan Pamuk dikkat çeken isimlerdir. Ahmet Ümit ile 2019 yılında anlaşılan banka bünyesindeki yayınevi bu başarılı grafiğini önümüzdeki yıllarda da sürdürme adına önemli bir atılım yapmıştır. Bu iki yayınevini Haruki Murakami, Elif Şafak, Zülfü Livaneli gibi başarılı yazarları bünyesinde barındıran bir diğer yayınevi takip etmektedir.

Analiz sonuçları bir arada değerlendirildiğinde, önümüzdeki yıllarda yerli ve yabancı roman arasındaki rekabetin süreceği görülmektedir. Özellikle sosyal medyanın rolü ve global film endüstrisinin yatırımlarıyla yabancı romanların üstünlüğünü sürdüreceği öngörülebilir. Yayınevleri bünyesindeki son derece başarılı çevirmenlerin varlığı da Türkiye’de yabancı romanlar açısından ciddi bir avantaj sağlamaktadır. Dini kitaplarla ilgili özellikle genç kuşak ekseninde bir araştırma yapılması gerektiği anlaşılmakta olup, aksi takdirde bu kategorinin çok satanlar listelerine girmesi çok olası görünmemektedir. Türkiye’de siyaset günlük hayatın bir parçası olmaya devam ettikçe politika kitapları güncel kalmaya ve satılmaya devam edecektir. Kişisel gelişim ve polisiye kitapları sabit kitleleri ile gelecekte de pazar paylarını koruyacaktır. İlber Ortaylı, Emrah Safa Gürkan gibi isimler genç nesillerde popülerliğini koruduğu ayrıca ana akım medyada tarihi diziler yer bulduğu sürece tarih kitapları da ilgi görmeye devam edecektir. Basımları sıklıkla yenilenen klasikler her dönemde okuyucuya hitap edebileceklerdir. Şiir ve öykü kategorileri ise ne yazık ki gelecekte çok fazla satış yapması beklenmeyen kategoriler olarak öne çıkmaktadırlar.

Sonuç

2014 - 2019 yılları arasında Türkiye’de yer alan üç büyük online kitap satış sitesi üzerinden elde edilen verilerle yapılan bu çalışmada, veri madenciliği tekniği kullanılarak kümeleme yöntemiyle hangi yılda, hangi yazarın ve o yazara ait hangi kitap kategorisinin çok veya az sattığı bilgilerine ulaşılmıştır. Ayrıca, en çok satan yayınevleri de analiz kapsamında değerlendirilmiştir.

Çalışmanın bulgularına göre, yerli ve yabancı romanlar en çok satılan kategori olmuştur. İçinde barındırdığı geniş yazar ve konu yelpazesi açısından özellikle klasikler anlamında yayınevlerinin en çok tercih ettiği kategorinin roman olması son derece doğaldır. Pek çok farklı yayınevini aynı klasiği basabilmesinin yasal olarak önünün açık olması da bu türü popüler bir konuma

getirmektedir. Bunların dışında, bu tarz klasik kitapların Milli Eğitim Bakanlığı’nun temel eserler olarak belirlediği listede yer alıyor olması da satış grafiğine olumlu yansımaktadır. Bilinirliği yüksek olan yayınevlerinin kapak görseli, kağıt kalitesi, doğru kapak dizaynı ve kaliteli çeviri gibi unsurları da dikkate alarak bu tür kitapları basıp satışa sunuyor olmaları gerek prestij gerekse ciro ve karlılık anlamında kendilerine olumlu bir şekilde geri dönecektir. Özellikle içinde bulunduğumuz pandemi sürecinde online kitap satışları yayınevlerinin finansal performansı açısından olumlu bir gelişme olarak kaydedilebilecektir. Bu anlamda yükseliş trendi içerisinde olan tüm kategoriler finansal performanslarını yükseltmek isteyen yayınevleri açısından son derece doğru tercihler olarak yerlerini alacaklardır. Online kanallarda sitenin tüketici açısından kolaylık sağlayan teknolojik altyapısı ve dağıtım kanalları ile tedarik zincirinin kalitesi rekabette kimin öne çıkacağına başlıca belirleyicileri olacaktır.

Bu çalışmanın en önemli özelliği, dijital çağda zorlu bir rekabeti içinde barındıran online kitap satışlarını veri madenciliği tekniği ile incelemektir. Ulaşılan sonuçlarla, basımı ve satışı yapılacak olan kitaplara yönelik yazar ve yayınevlerine doğru bir bilgilendirme sağlanması ve hem fikir aşamasındaki yazarlara hem de işin maliyetini üstlenecek olan yayınevlerine stratejilerini doğru kurgulama fırsatı sunmak hedeflenmektedir.

Araştırmanın sadece üç yayınevi ile yapılmış olması, metodolojide de değinildiği üzere, bu çalışmanın bir kısıtı olup buna bağlı olarak sonuçların genelle yayılmasının çok da doğru olmayacağına altı çizilmelidir. Ancak, bu yayınevi sitelerinin gerek yüksek ciroları gerekse müşteri portföyü açısından zengin bir altyapıya sahip oluşları Türkiye geneli hakkında bir fikir edinme açısından kayda değerdir.

İleriki çalışmalara ışık tutması açısından, veri madenciliği içerisindeki farklı modeller kullanılarak kitaplara ilişkin bu tür araştırmalar genişletilebilir. Ayrıca, kitap kategorileri ile finansal performansı belirleyici kalemler arasındaki ilişkilerin irdeleneceği bir diğer çalışma da sektör açısından son derece katma değer yaratıcı olabilecektir.

EXTENDED ABSTRACT

**The Analysis of Online Book Selling in Turkey by
Data Mining**

*

Meltem Ulsan Polat – Mehmetcan Babaoğlu
Marmara University, Mercedes – Benz Otomotiv

Online sales branches take up a big space in today's technological world and their emphasis on our daily lives increase at breakneck speed. The huge rivalry of book-selling web-sites, enabling wide product range access, increase the attractiveness of online book sales. During current pandemic period, online book sales has been one of the major sectors in terms of interest and accordingly profitability. Online book shopping has been triggered during the lockdown periods, since it was one of the best options at home-stay. The study has a remarkable significance analyzing reading habit of Turkish society and planning sound strategies considering readers' choices. Furthermore, it could also be perceived as a guidance for publishers to develop their future plans regarding cooperation with the authors and preparing their annual budgets for the upcoming period. Within the scope of this study, data about best-selling books from 2014 to 2019 have been gathered from the most trustable and popular three web-sites. The mentioned data comprise variables such as title, author, genre and publisher. The quantity of the books is 508. The data have been analyzed using Clustering - K Means algorithm, as one of the data mining methods, by RapidMiner Program. In Clustering, different data objects are initially classified in several groups as similar objects then data are summarized and labeled. Clustering method has different fields of usage such as marketing, financial performance and banking. K-Means is a clustering algorithm. The algorithm targets to partition observations into k clusters in a way that each observation refers to the cluster with the closest mean. Since the data were annually - based, it has also been taken necessary to use K- Means algorithm. RapidMiner, software program used in the study, supports all the stages of data mining process, like data preparation, visualization of findings, data validation and optimization. The findings within the scope of the variables introduced the leading position of foreign novels. Russian literature and fantastic

fiction are the most preferred novels in this category. Since many of the books have been filmed, it also has increased their popularity and recognition. Foreign novels are followed by native Turkish novels. Sabahattin Ali, Zulfu Livaneli and Ahmet Umit are the most popular authors of native Turkish books. One of the most remarkable findings of the study is the decline of religion-based thematic books. A study regarding the detailed content of these books and their meeting readers’ needs by considering especially the young generation could be conducted for future surveys. This could also have a positive contribution to the authors and the publisher of this category. Another interesting finding is the remarkable increase of history books. İlber Ortaylı is the most popular of the of this category. Self-help books indicate a stable trend, since they have a proper group of readers. The sale of children’s books increase as well however, it is not possible to propound that reading habit among children goes up just by considering the mentioned data. As another result, it is seen that the sale of short story books plunged. The reason could be the monthly-published literature magazines. Poetry books are the another declining category, since readers, especially the young generation, prefer to reach poems via social media platforms. In line with the popularity of policy-based discussion programs on TV, policy-based books started to attract attention. In terms of authors, Franz Kafka, Elif Shafak, Iskender Pala and Paulo Coelho are very popular. As the final finding, publishers have been analyzed and it is revealed that two big publishers are leading. One has been founded by a famous author in the 80’s and the other is the subsidiary of one of the biggest banks in Turkey. Furthermore, one is famous with its foreign novels while the other is leading with native Turkish novels. By considering the overall data, it is seen that the competition between foreign and native Turkish novels will continue. The results have been deeply evaluated particularly to guide authors at the idea stage and to determine a road map for publishers while bearing the costs. As the restriction of the study, it is important to emphasize that the variables of only three publishers have been chosen. However, these web-sites have huge amounts of sales and they have a wide range of customer portfolio. Therefore, the findings are remarkable to give a general idea of Turkish society. For further studies, different methods in data mining could be used to analyze book-sale-related studies. Furthermore, financial-performance-based studies could also have an added value for the academic literature and for the sector, as well.

Kaynakça / References

- Akça, M. (2014). *K-Means kümeleme algoritması nedir?* 08.04.2020 tarihinde <http://mustafaakca.com/k-means-kumeleme-algoritmasi/> adresinden erişilmiştir.
- Aksan, C. E. (2019). *RapidMiner nedir?* 08.04.2020 tarihinde <https://ceaksan.com/tr/rapidminer-nedir/> adresinden erişilmiştir.
- Arı, E., Özköse, H., Doğan, A., ve Calp, M. (2016). İstanbul Borsası'nda işlem gören firmaların finansal performanslarının kümeleme analizi ile değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 9(1), 33-0. <http://dx.doi.org.tr/10.17671/btd.55726>
- Börühan, G., Yumurtacı Hüseyinoğlu, İ.Ö., ve Ersoy, P. (2017). Perakendecilerin e-ticaret ve tüketici algısına yönelik bir araştırma. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 20(37), 197-218.
- Çelik, M. (2009). *Veri madenciliğinde kullanılan sınıflandırma yöntemleri ve bir uygulama*. Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü. <http://acikerisim.istanbul.edu.tr/bitstream/handle/123456789/28001/45671.pdf?sequence=1&isAllowed=y> adresinden erişilmiştir.
- Emel, G.G., ve Taşkın, Ç. (2005). Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2), Aralık, 221-239.
- Kalikov, A. (2006). *Veri madenciliği ve bir e-ticaret uygulaması*. Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Kara, H. (2018, 17 Haziran). *Güvenilir kitap satın alma siteleri*. <https://www.hknkr.com/guvenilir-kitap-satin-alma-siteleri.html> adresinden erişilmiştir.
- Lin, T.Y., Ohsuga, S., Liau, C.J., ve Hu, X. (2005). *Foundations and novel approaches in Data Mining, studies in computational intelligence*, 9. Springer.
- Nettleton D., Baeza-Yates R., ve Marcos M.C. (2013). Analysis of the user queries of an e-Commerce bookstore in terms of the LCC catalogue and publishers. *Information Research*, 18(4), Paper 596, December. <http://InformationR.net/ir/18-4/paper596.html>
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Dergisi* 3, 65-82.
- Özgür, S. (2012). *Elektronik ortamda kitap satışı veritabanı yapısı örnek bir uygulama*. Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü. <https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=QVPIT7WD6kjUIE-RiihXnQ&no=hljioIwIYUKs WcB Qzj Tg> adresinden erişilmiştir.
- Özkan, Y. (2008). *Veri madenciliği yöntemleri*. Papatya Bilim Yayınları.

- Rowe, A. (2019, 5 Ocak). *7 Publishing insights revealed by last year’s top 100 bestselling books*. *Forbes*. <https://www.forbes.com/sites/adamrowe1/2019/01/05/publishing-insights-revealed-by-last-years-top-100-best-selling-books/#5d671fe86913> adresinden erişilmiştir.
- Sarman, G. (2011). Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: K-Means ve K-Medoids kümeleme algoritmalarının karşılaştırılması. *Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü Dergisi*, 15(3), 192-202.
- Savaş, S., Topaloğlu, N., ve Yılmaz, M. (2012). Veri madenciliği ve Türkiye’deki uygulama. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11 (21), 1-23.
- Sharma, S.K., Chakraborti, S., & Jha, T. (2019). Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach. *Information Systems and E-Business Management, Springer*, 17(2), 261–284. <https://doi.org/10.1007/s10257-019-00438-3>
- T.C. Kültür ve Turizm Bakanlığı Kütüphaneler ve Yayımlar Genel Müdürlüğü. (2011). *Türkiye okuma kültürü haritası*. <https://kygm.ktb.gov.tr/TR-36/turkiye-okuma-kulturu-haritasi-arastirmasi-sonuclari-ac.html>
- Türkiye Yayıncılar Birliği. (2020). *Türkiye kitap pazarı raporu*. <http://turkaybir.org.tr/2019-turkiye-kitap-pazarı-raporu/#.Xue5rUUzaM8> adresinden erişilmiştir.
- Yıldız, M., ve Şeker, Ş.E. (2016). Veri madenciliği araçları. *YBS Ansiklopedi*, 3(4), Aralık, 10-20.
- Yücesoy, B., Wang, X., Huang, J., ve Barabási, A. L. (2018). Success in books: a big data approach to bestsellers. *EPJ Data Science* 7. <https://doi.org/10.1140/epjds/s13688-018-0135-y>

Kaynakça Bilgisi / Citation Information

Polat Ulusan, M. ve Babaoğlu, M. (2021). Türkiye’de online kitap satışlarının veri madenciliği yöntemiyle analizi. *OPUS–Uluslararası Toplum Araştırmaları Dergisi*, 17(36), 2794-2815. DOI: 10.26466/opus.793764